NeuroImage 224 (2021) 117427

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

Modelling the effects of transcranial alternating current stimulation on the neural encoding of speech in noise

Mikolaj Kegler, Tobias Reichenbach*

Department of Bioengineering and Centre for Neurotechnology, Imperial College London, South Kensington Campus, SW7 2BU London, United Kingdom

ARTICLE INFO

Keywords:
Speech processing
Transcranial alternating current stimulation
Spiking neural networks
Computational modelling

ABSTRACT

Transcranial alternating current stimulation (tACS) can non-invasively modulate neuronal activity in the cerebral cortex, in particular at the frequency of the applied stimulation. Such modulation can matter for speech processing, since the latter involves the tracking of slow amplitude fluctuations in speech by cortical activity. tACS with a current signal that follows the envelope of a speech stimulus has indeed been found to influence the cortical tracking and to modulate the comprehension of the speech in background noise. However, how exactly tACS influences the speech-related cortical activity, and how it causes the observed effects on speech comprehension, remains poorly understood. A computational model for cortical speech processing in a biophysically plausible spiking neural network has recently been proposed. Here we extended the model to investigate the effects of different types of stimulation waveforms, similar to those previously applied in experimental studies, on the processing of speech in noise. We assessed in particular how well speech could be decoded from the neural network activity when paired with the exogenous stimulation. We found that, in the absence of current stimulation, the speech-in-noise decoding accuracy was comparable to the comprehension of speech in background noise of human listeners. We further found that current stimulation could alter the speech decoding accuracy by a few percent, comparable to the effects of tACS on speech-in-noise comprehension. Our simulations further allowed us to identify the parameters for the stimulation waveforms that yielded the largest enhancement of speech-in-noise encoding. Our model thereby provides insight into the potential neural mechanisms by which weak alternating current stimulation may influence speech comprehension and allows to screen a large range of stimulation waveforms for their effect on speech processing.

1. Introduction

Naturalistic listening environments are often noisy. Talking to a friend in a busy pub or restaurant, for instance, means that we need to ignore other distracting sounds around us. However, humans excel at this challenging task: we can still understand speech even when the background noise becomes louder than the target signal itself (Hutcherson et al., 1979; Drullmana, 1995; Soli and Wong, 2008; Anderson and Kraus, 2010).

This remarkable performance partly involves the tracking of amplitude fluctuations in speech by cortical activity (Hickok and Poeppel, 2007; Morillon et al., 2012; Mesgarani et al., 2014; Han et al., 2019). In particular, the neural oscillations in the delta (1 - 4 Hz) and theta (4 - 8 Hz) frequency ranges become correlated with the acoustic envelope of a speech stimulus (Lalor and Foxe, 2010; Kubanek et al., 2013; Molinaro and Lizarazu, 2018; Brodbeck and Simon, 2020). They can thereby track the rhythm set by words (in the delta range) and by syllables (in the theta range). When a speech stimulus is obscured by background noise, such as a competing speaker, this low-frequency cortical tracking can predict speech discrimination performance (Luo and Poeppel, 2007), selective attention (Zion Golumbic et al., 2013; O'Sullivan et al., 2015, 2017), speech intelligibility (Vanthornhout et al., 2018; Lesenfants et al., 2019) and comprehension (Etard and Reichenbach, 2019; Iotzov and Parra, 2019). The delta and theta frequency band thereby play different roles: cortical tracking in the theta band is linked to lower-level acoustic processing of the speech stimulus, while delta-band tracking can inform on higher-level aspects such as the processing of semantic and syntactic information (Ding et al., 2016; Broderick et al., 2018; Etard and Reichenbach, 2019).

Neural tracking of speech features has also been demonstrated in a higher frequency band, the gamma band. It contains activity above 25 Hz and can encode phonemes, the basic units of speech (Shamir et al., 2009; Gross et al., 2013). A recent hypothesis postulates that speech processing occurs through a cross-frequency coupling of cortical oscillations (Giraud and Poeppel, 2012; Gross et al., 2013). According to this hypothesis, the cortical activity in the theta band parses speech into

* Corresponding author.
E-mail address: reichenbach@imperial.ac.uk (T. Reichenbach).

https://doi.org/10.1016/j.neuroimage.2020.117427
Received 8 June 2020; Received in revised form 11 September 2020; Accepted 1 October 2020
Available online 7 October 2020

smaller units, presumably syllables (Giraud et al., 2007; Ghitza, 2011). The theta activity then modulates the cortical responses in the gamma range, thus providing temporal frames for the phonemic encoding.

Transcranial alternating current stimulation (tACS) provides a non-invasive means to influence cortical activity in humans, in particular at the frequency of the stimulation (Zaehle et al., 2010; Reato et al., 2013; Helfrich et al., 2014; Ruhnau et al., 2016; Krause et al., 2019). Sinewave tACS combined with the rhythmic presentation of a speech stimulus has indeed been shown to affect the cortical responses to speech (Zoefel et al., 2018). Moreover, tACS with the speech envelope impacts behaviour as well: the comprehension of speech in noise can be modulated through concurrent neurostimulation (Wilsch et al., 2018; Kadir et al., 2020; Keshavarzi and Reichenbach, 2020; Keshavarzi et al., 2020). The modulation is modest, up to a few percent in the comprehension scores. It results from the theta but not the delta portion of the speech envelope, indicating that the stimulation may act on the syllable parsing (Keshavarzi et al., 2020). Moreover, the current stimulation in the theta band can boost the comprehension of speech in background noise beyond that observed during sham stimulation (Keshavarzi and Reichenbach, 2020; Keshavarzi et al., 2020).

The experimental data regarding the effect of tACS with the speech envelope on speech comprehension show, however, certain inconsistencies. Two key variables that have been explored when applying tACS simultaneous to speech in noise are the delay between the current waveform and the speech envelope, as well as a potential phase shift between these two signals. Some studies found that the value of the stimulation parameter, either of the delay or of the phase shift, that yielded the highest speech comprehension varied considerably between subjects (Riecke et al., 2018; Wilsch et al., 2018). These results suggest that the current stimulation acts on a cortical source that is highly variable from subject to subject. In contrast, other studies found that the optimal delay and phase shift of the current waveform with respect to the speech signal were similar across different study participants (Kadir et al., 2020; Keshavarzi and Reichenbach, 2020; Keshavarzi et al., 2020). The inconsistencies between these different investigations provide additional motivation for better understanding the functional mechanisms by which tACS influences speech comprehension.

Computational modelling offers a promising route to investigate the effects of non-invasive brain stimulation (Fröhlich and Schmidt, 2013; Bestmann et al., 2015; Bonaiuto and Bestmann, 2015; Fröhlich, 2015; Fröhlich et al., 2015). Well-established finite-element models that are based on structural imaging data are, for instance, used to estimate the distribution of electrical current in the brain (Datta et al., 2009; Huang et al., 2019). They allow to optimize the placement of electrodes on the scalp and can explain some inter-subject variability (Huang and Parra, 2019; Kasten et al., 2019). They do, however, not provide information on the functional mechanisms by which the current stimulation influences the neural network activity underlying the behavioural effects.

The functional influence of current stimulation can be addressed through biophysically-plausible spiking neural network models combined with a model of how each neuron's activity is affected by a weak current (Reato et al., 2010; Ali et al., 2013; Herrmann et al., 2016; Cakan and Obermayer, 2020). Recent effort in this direction has, for instance, uncovered that tACS can act on cortical oscillations through periodic forcing (Fröhlich and McCormick, 2010; Reato et al., 2010; Herrmann et al., 2016; Cakan and Obermayer, 2020) as known from other nonlinear dynamical systems (Pikovsky et al., 2001). However, the functional mechanisms of current stimulation in relation to sensory processing have not yet been investigated computationally.

Here, we introduce a framework for modelling the effects of external electrical stimulation, similar to tACS, on the neural encoding of speech in background noise. Our computational work is based on a recently introduced model of speech encoding through coupled cortical oscillations in the theta and in the gamma frequency ranges (Hyafil et al., 2015). We show that the model can be used to describe the encoding of speech in background noise. We then extend it to include the effects of alternating current stimulation and employ it to investigate the mechanism by which current stimulation affects the speech encoding.

2. Methods

2.1. Computational model of speech encoding

We employed a computational model for speech encoding in a spiking neural network (Hyafil et al., 2015). The model consisted of two modules of spiking neurons that generated endogenous oscillations in the theta (4 - 8 Hz) and in the gamma (25 - 40 Hz) frequency ranges (Fig. 1A). The gamma oscillations resulted from a Pyramidal Interneuron Gamma (PIN-G) module (Jadi and Sejnowski, 2014). In this well established and experimentally validated model, a group of excitatory neurons and another group of inhibitory neurons are reciprocally connected to each other to generate oscillations (Brosch et al., 2002; Cardin et al., 2009; Sohal et al., 2009; Ray and Maunsell, 2011). Since the mechanisms of the neural activity in the theta frequency range remain unknown (Ainsworth et al., 2011), the theta-generating module was designed analogously to the gamma module, but with adjusted parameters such as slower time scales, and was referred to as PIN-TH model.

The spiking neural network model contained 84 leaky integrate-and-fire neurons of four distinct types: gamma excitatory neurons (*Ge*, $N_{Ge} = 32$ cells), gamma inhibitory neurons (*Gi*, $N_{Gi} = 32$ cells), theta excitatory neurons (*Te*, $N_{Te} = 10$ cells) and theta inhibitory neurons (*Ti*, $N_{Ti} = 10$ cells). The first two types of neurons formed the PIN-G module, and the second two types belonged to the PIN-TH module.

The temporal evolution of the membrane potential V_i of neuron i is described by the following equation:

$$C\frac{dV_i}{dt} = g_L\left(V_L - V_i\right) + I_i^{SYN} + I_i^{INP} + I_i^{EXT} + I_i^{DC} + \eta, \tag{1}$$

in which C is the capacitance of the cellular membrane, g_L and V_L are the conductance and the reversal potential of the leak current; I_i^{SYN}, I_i^{INP}, I_i^{EXT} and I_i^{DC} are the synaptic, stimulus-induced, exogenous and constant currents delivered to the cell, and η is a Gaussian noise with variance σ_i. When the membrane potential of the ith neuron reached the threshold V_{THR} a spike was generated and V_i returned to the reset potential V_{RESET}.

The dynamics of synaptic currents between neurons were modelled as follows:

$$\frac{dx_{ij}^R}{dt} = -\frac{x_{ij}^R}{\tau_j^R} + \delta\left(t - t_j^{SPK}\right), \tag{2}$$

$$\frac{ds_{ij}}{dt} = \frac{x_{ij}^R - s_{ij}}{\tau_j^D}, \tag{3}$$

where s_{ij}, x_{ij}^R are activation variables of the synapse at neuron i for a connection coming from neuron j, $\delta(t - t_j^{SPK})$ indicates a spike generation in the presynaptic neuron at the time t_j^{SPK}, and τ_j^R, τ_j^D are time constants that describe the rise and decay of the activation from neuron j, respectively. The synaptic current I_i^{SYN} is then the sum of all synaptic inputs to neuron i from the remaining cells:

$$I_i^{SYN}(t) = \sum_j g_{ij} s_{ij}(t)\left(V_j^{SYN} - V_i(t)\right), \tag{4}$$

where g_{ij} is the synaptic conductance of the synapse from neuron j to i, and V_j^{SYN} is the equilibrium potential of the presynaptic neuron j.

Because we only modelled a small and local neural network, we employed all-to-all connections between the different neurons of each subtype. The PIN-G and PIN-TH modules were then created by reciprocally coupling the corresponding excitatory and inhibitory neurons, that is,

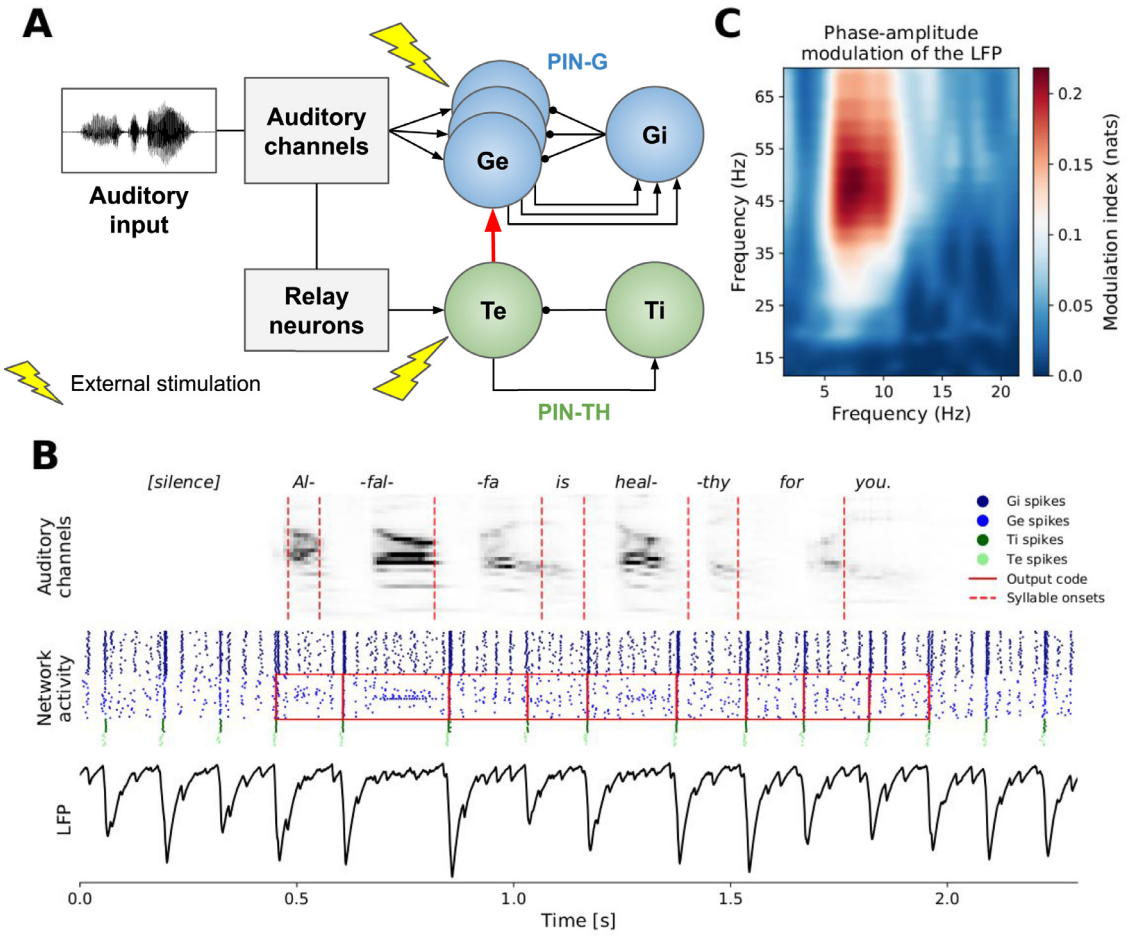

Fig. 1. Architecture of the spiking neural network and its dynamics. (A) Network architecture. A PIN-TH module (green) consisted of 10 excitatory neurons (*Te*) and 10 inhibitory neurons (*Ti*) to generate self-sustained oscillations in the theta frequency band. Analogously, a PIN-G module (blue) with 32 excitatory cells (*Ge*) and 32 inhibitory cells (*Gi*) produced faster gamma-range activity. Both modules were coupled unidirectionally through all-to-all connections from the *Te* to the *Ge* cells. The auditory input to the model was firstly decomposed into 32 frequency-specific auditory channels, using a model of the auditory periphery. The resulting signals were projected to *Ge* neurons. They were also convolved with a spectrotemporal filter that mimicked the action of relay neurons and then fed into the *Te* neurons. The application of transcranial current stimulation (yellow) was simulated as a current injection to all excitatory cells in the model. **(B)** The network's response to the example sentence *'Alfalfa is healthy for you'*, preceded by silence. The model of the auditory periphery decomposed the sound into 32 auditory channels (top). The resulting neural spikes from the theta module (middle, green) allowed to infer syllable boundaries, and to group the neural output of the gamma module (middle, red boxes) according to the individual syllables, enabling the decoding of the syllable identity. The local field potential (LFP, bottom) followed as the sum of the synaptic currents delivered to the excitatory neurons. **(C)** The coupling from the theta module to the gamma module resulted in phase-amplitude modulation. In particular, the phase-amplitude modulation index was high for phases in the theta range, around 5 – 12 Hz, and for amplitudes between 35 - 70 Hz, in the gamma range.

those of type *Ge* and *Gi* respectively those of type *Te* and *Ti*. In addition, the *Ti* neurons were all-to-all connected to facilitate sparse synchronous spiking within this population. The cross-frequency coupling in the model was implemented by connecting the PIN-G module to the PIN-TH module through unidirectional all-to-all connections from the *Te* to the *Ge* neurons.

The values of the model parameters were obtained from the study that introduced the model (Hyafil et al., 2015), and are listed in Table 1. Eqs. (1)–(4) were solved numerically using the Euler method with a time step of 10 μs. The local field potential (LFP) was obtained by summing the absolute values of all synaptic currents delivered to the excitatory cells *Ge* and *Te* in the network (Mazzoni et al., 2008).

2.2. Simulation of alternating current stimulation in the model

Following recent computational models for the effects of tACS on neural oscillations, we simulated the neurostimulation as a current injected to all excitatory neurons in the network (Ali et al., 2013; Herrmann et al., 2016; Negahbani et al., 2018) (Fig. 1A, yellow). Experimental evidence suggests indeed that pyramidal neurons, the excitatory ones, are significantly more susceptible to external electric fields than the inhibitory interneurons (Radman et al., 2009).

To calibrate the intensity of the exogenous stimulation, a constant stimulation current I^{ext} was applied to an isolated *Ge* pyramidal neuron. Specifically, the synaptic current I^{SYN}, the stimulus input current I^{INP}, as well as the constant current I^{DC} were all set to 0 with the remaining parameter values unchanged. The external current was applied 10 s after the start of a simulation, for a duration of 10 s. Its intensity was varied from 0.01 pA to 1 pA in steps of 0.01 pA. For each intensity of the external current, we ran 100 simulations. We thereby identified the spiking threshold of an isolated *Ge* neuron as 0.71 pA. This intensity of stimulation led, just below the spiking threshold, to an average membrane depolarization over 7 mV, comparable to the levels observed in previous computational models for the effects of tACS (Negahbani et al., 2018). Since non-invasive transcranial electrical stimulation in humans is not powerful enough to directly cause spiking in cortical neurons, in the following simulations we considered subthreshold stimulation at three intensities: 0.1 pA, 0.2 pA and 0.5 pA. These led to an average membrane depolarization of 1 mV, 2 mV and 5 mV, respectively.

Table 1
Model parameters.

Parameter	Description	Value
Neuron model		
C	Cell membrane capacitance	1 pF
V_{THR}	Spiking threshold	-40 mV
V_{RESET}	Resting potential	-87 mV
V_L	Equilibrium potential of leak	-67 mV
V_E^{SYN}	Equilibrium potential of excitatory neurons	0 mV
V_I^{SYN}	Equilibrium potential of inhibitory neurons	-80 mV
PIN-G network		
g_{LE}, g_{LI}	Leak conductance in *Ge*, *Gi* neurons	0.1 nS
τ_{Ge}^R	Synaptic rise constant of *Ge* neurons	0.2 ms
τ_{Gi}^R	Synaptic rise constant of *Gi* neurons	0.5 ms
τ_{Ge}^D	Synaptic decay constant of *Ge* neurons	2 ms
τ_{Gi}^D	Synaptic decay constant of *Gi* neurons	20 ms
I_{Ge}^{DC}	Constant current delivered to *Ge* neurons	3 pA
I_{Gi}^{DC}	Constant current delivered to *Gi* neurons	1 pA
σ_{Ge}, σ_{Gi}	Variance of the noise term in *Ge*, *Gi* neurons	2.028 pA$\cdot\sqrt{ms}$
PIN-TH network		
g_{LE}	Leak conductance in *Te* neurons	0.0264 nS
g_{LI}	Leak conductance in *Ti* neurons	0.1 nS
τ_{Te}^R	Synaptic rise constant of *Te* neurons	4 ms
τ_{Ti}^R	Synaptic rise constant of *Ti* neurons	5 ms
τ_{Te}^D	Synaptic decay constant of *Te* neurons	24.3150 ms
τ_{Ti}^D	Synaptic decay constant of *Ti* neurons	30.3575 ms
I_{Te}^{DC}	Constant current delivered to *Te* neurons	1.25 pA
I_{Ti}^{DC}	Constant current delivered to *Ti* neurons	0.0851 pA
σ_{Te}	Variance of the noise term in *Te* neurons	0.282 pA$\cdot\sqrt{ms}$
σ_{Ti}	Variance of the noise term in *Ti* neurons	2.028 pA$\cdot\sqrt{ms}$
Connectivity		
$g_{Ge,\ Gi}$	*Gi* → *Ge* connectivity	$5/N_{Gi}$ nS
$g_{Gi,\ Ge}$	*Ge* → *Gi* synaptic conductance strength	$10/N_{Ge}$ nS
$g_{Ge,\ Te}$	*Te* → *Ge* synaptic conductance strength	$1/N_{Te}$ nS
$g_{Te,\ Ti}$	*Ti* → *Te* synaptic conductance strength	$2.07/N_{Ti}$ nS
$g_{Ti,\ Te}$	*Te* → *Ti* synaptic conductance strength	$6.66/N_{Te}$ nS
$g_{Ti,\ Ti}$	*Ti* → *Ti* synaptic conductance strength	$4.32/N_{Ti}$ nS

2.3. Auditory stimuli and network simulations

Spoken English sentences from the TIMIT dataset (Garofolo et al., 1993) at a sound-pressure level of 76 dB SPL were used as input to the neural network model. To investigate speech-in-noise encoding in the model, we chose a random subset of 100 sentences. We added four-talker babble noise to each sentence at signal-to-noise ratios (SNRs) that ranged from -25 to 25 dB, in steps of 5 dB. The SNR was thereby determined from the ratio of the root-mean-square amplitudes of the signal and of the background noise.

For each SNR and each sentence, we simulated the neural network response 100 times. Because the theta module generated intrinsic oscillatory activity, we wanted to prevent an accidental alignment between this theta activity and the onset of the speech. Each sentence was therefore preceded by a silent period whose duration varied randomly between 380 ms and 550 ms. Each simulation was terminated 100 ms after the end of the presented sentence.

To investigate the effect of the neural coupling between the PIN-TH module and the PIN-G module, we employed a simpler simulation setup: we computed the LFP in response to the exemplary sentence '*Alfalfa is healthy for you.*'. The model responses were simulated 30 times, and in each simulation the sentence was preceded by a random period of silence that ranged from 500 ms to 1,000 ms.

2.4. Input of the acoustic signal to the neural network

Following the previously introduced model for speech processing through a coupled PIN-TH and PIN-G modules (Hyafil et al., 2015), the auditory input was processed through a model of the auditory periphery (Chi et al., 2005). This model firstly decomposed the auditory stimulus through a cochlear filter bank into 128 channels. The signals in the different channels were then subjected to nonlinear transformations that reflected neural processing in the auditory nerve and the subcortical nuclei. First, mimicking the action of hair cells, the filtered signals were high-pass filtered, nonlinearly compressed and then low-pass filtered (Yang et al., 1992). Second, a first order derivative across frequency channels was taken, followed by a half-way rectification, which reflected the lateral inhibition in the cochlear nucleus (Shamma, 1989). Third, the signal in each channel was integrated over a short temporal duration of 8 ms, reflecting the decay of temporal precision in the midbrain. The obtained signals were interpreted as currents measured in pA, and approximated the tonotopically organized input to the primary auditory cortex.

The auditory stimuli processed through the model of the auditory periphery were projected to both the PIN-G and the PIN-TH module. First, regarding the PIN-G module, each of the 32 *Ge* neurons received input from one auditory channel, in a tonotopic fashion. To this end, the number of auditory channels was reduced to 32 by selecting every fourth auditory channel from all 128 available.

Second, the sound stimuli were used as input to the slower PIN-TH module as well. In particular, the *Te* neurons were stimulated in a way that tracked syllable onsets as faithfully as possible. To this end, the *Te* neurons received an input current $Y(t)$ that was the convolution of the 32 auditory channels described above with a spectrotemporal filter at auditory channel c and delay τ:

$$Y(t) = \sum_{c=1}^{32} \sum_{i=1}^{6} B(c,\tau) X\left(c, t-\tau_i\right), \tag{5}$$

in which X is the signal in the auditory channel c at time $t-\tau_i$. The 6 temporal delays τ_i were uniformly distributed between -50 ms and 0 ms. The convolution of the auditory input with the filter B modelled the effect of a population of relay neurons with delays of up to 50 ms, and with weights that represent the strength of synaptic connections (Pillow et al., 2008). Unlike the tonotopically organized *Ge* neurons, all *Te* cells received the same current input $Y(t)$.

The spectrotemporal filter B was computed from 1,000 randomly chosen sentences from the TIMIT corpus to optimize the predictions of the syllable onsets (Hyafil et al., 2015). These sentences differed from the ones that were used for subsequent investigations of speech coding in the neural network. The audio signals were preceded by a silent part whose duration varied randomly between 500 ms and 1,000 ms. The signals were processed by the model of the auditory periphery, downsampled to 100 Hz and concatenated to obtain the signals X.

The onsets of syllables were obtained from the TIMIT transcription, and were used to compute a binary vector. The syllable onsets in this vector were shifted forward by 20 ms such that they occurred after the actual onsets. The filter coefficients B were then computed through sparse bilinear logistic regression to predict this syllable vector, with the syllable onset vector replacing the current input $Y(t)$ in Eq. (5) (Shi et al., 2014; Adam and Hyafil, 2020).

2.5. Stimulation waveform design

We explored stimulation waveforms that were based on the envelope of the speech stimuli (Fig. 2). The envelope of a sentence was computed by determining the analytic representation of the speech signal using the Hilbert transform, and by calculating its absolute value. The obtained signal was further band-pass filtered between 1 - 4 Hz, between 4 - 8 Hz, or between 1 - 20 Hz, yielding the delta portion of the speech envelope, the theta portion of the envelope, or the broadband envelope, respectively (2^{nd} order, zero-phase Butterworth bandpass filter).

We then shifted the obtained envelopes by six different phases, ranging from 0° to 300°, in steps of 60°. In particular, the shift of an envelope $e(t)$ by a phase ϕ was implemented through the Hilbert transform $H[e(t)]$, yielding the analytical representation $E(t)$ of the envelope:

$$E(t) = e(t) + i \cdot H[e(t)], \tag{6}$$

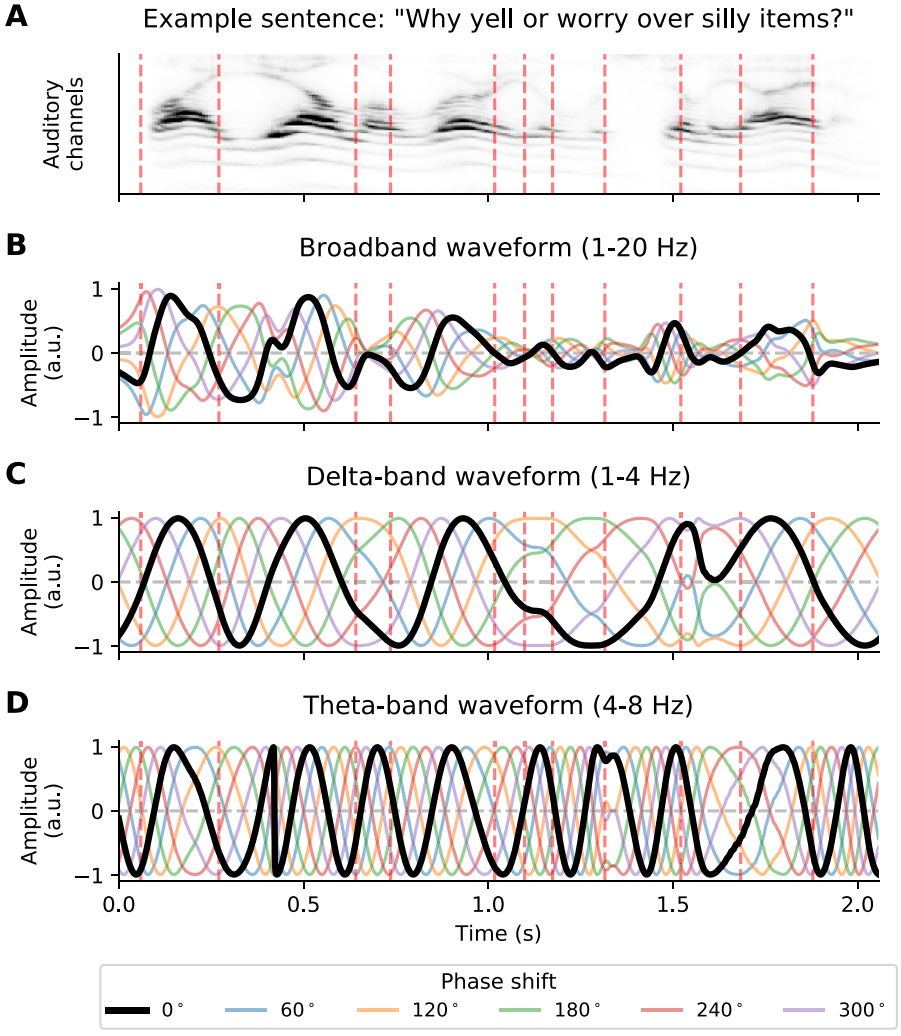

Fig. 2. Envelope-shaped stimulation waveforms. Stimulation waveforms derived from the exemplary TIMIT sentence *'Why yell or worry over silly items?'*. **(A)** Exemplary sentence decomposed into 128 auditory channels and its syllable boundaries obtained from the TIMIT's phonetic transcription (dashed red lines). **(B)** - **(D)** Waveforms for the neurostimulation were derived from the speech envelope, filtered into a broadband frequency range (**B**), into the delta range (**C**), or into the theta range (**D**). The waveforms in the delta and in the theta band were altered so that the maxima and minima occurred at the values of 1 and -1, respectively. All waveforms were then shifted by six different phases (coloured).

in which i denotes the imaginary unit. The phase-shifted envelope $e_\phi(t)$ then followed as

$$e_\phi(t) = |E(t)|\mathrm{Re}\left(e^{i\{\arg[E(t)]+2\pi\phi/360°\}}\right). \quad (7)$$

For the two narrowband stimulation signals, the ones that were filtered in the delta and in the theta ranges, we processed the waveforms further such that all the maxima had the same value, and that the minima had the opposite value. We recently employed such signals in an experimental investigation on the effects of current stimulation on speech comprehension (Keshavarzi and Reichenbach, 2020; Keshavarzi et al., 2020). To obtain these waveforms, the amplitude of the analytical envelope, $|E(t)|$, was set to 1 in Eq. (7).

For the broadband stimulation waveform, 1 - 20 Hz, we kept its original, non-fixed, amplitude, since this enabled comparison with previous experimental work (Wilsch et al., 2018; Kadir et al., 2020). In addition, processing these waveforms to achieve maxima and minima at equal amplitudes would have introduced major distortion to the signals.

Each phase-shifted envelope $e_\phi(t)$ was then normalized such that no value of the waveform either exceeded 1 or fell below -1. The neurostimulation was simulated in the model by multiplying a particular stimulation waveform by the desired stimulation intensity.

In order to investigate how the temporal alignment of the envelope-shaped stimulation waveform with the acoustic input influenced the speech processing, we employed stimulation waveforms without phase shift (i.e. with a phase shift of 0°) but with different temporal delays. We employed time lags ranging from -250 ms to 250 ms in steps of 50 ms step, with positive lags representing a stimulation waveform that preceded the acoustic stimulation.

2.6. Analysis of the phase-amplitude modulation

The spiking neural network was designed such that the phase of the theta oscillations influenced the amplitude of the gamma oscillations. To quantify this coupling, we computed the phase-amplitude modulation index from the LFP (Tort et al., 2010). In particular, we computed the LFP in response to the exemplary sentence '*Alfalfa is healthy for you.*'. The model responses were computed independently 30 times, and the sentence was each time preceded by a random period of silence that ranged from 500 to 1,000 ms. The simulated LFP was then downsampled to 1,000 Hz. It was further subjected to the complex Morlet wavelet transform with frequencies between 1 and 80 Hz, in steps of 0.1 Hz. For each frequency, the extracted amplitudes were binned into 18 bins according to their instantaneous phases. The phase-amplitude modulation index was computed as the Kullback–Leibler divergence (Kullback and Leibler, 1951) of the amplitude distribution across the phase bins from a uniform distribution.

2.7. Analysis of syllable parsing

The theta module PIN-TH produced only sparse spiking activity (Fig. 1). Spikes that occurred synchronously across different neurons emerged rhythmically in silence and followed syllable boundaries in response to speech. We accordingly employed the model to infer syllable

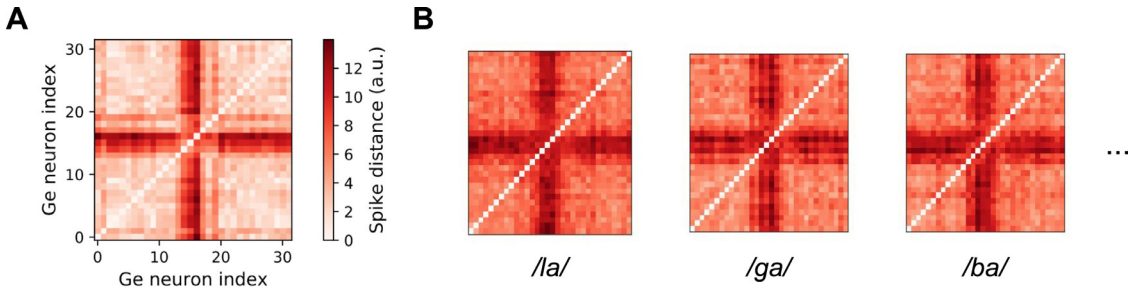

Fig. 3. Syllable decoding. **(A)** The identity of a syllable was decoded from the corresponding chunk of the neural response of the gamma module. To this end, the outputs of the 32 *Ge* neurons in that segment were characterized by their pairwise dissimilarity matrix. One such matrix was obtained for each syllable parsed by the PIN-TH module of the network (Fig. 1B, red boxes). **(B)** The pairwise dissimilarity matrices of the *Ge* neuronal responses differed for different syllables. To decode the identity of an unknown syllable from the neuronal response, its dissimilarity matrix was compared to the averaged dissimilarity matrices for the different syllables, obtained from simulations employing clean speech. The unknown syllable was then assigned the identity of the nearest clean-speech dissimilarity matrix.

onsets by detecting spike bursts. We thereby defined a spike burst as the spiking activity of at least two inhibitory neurons, which had sparser spiking activity than the excitatory neurons, within a 20 ms window. The precise timing of the syllable onset was assigned according to the maximal firing rate of the *Ti* neurons, computed using sliding 20-ms-long gaussian window with a standard deviation of 3 ms.

The performance of the resulting syllable parser was assessed by computing the distance, or dissimilarity, between the actual syllable boundaries and those inferred from the activity of the PIN-TH module. We thereby measured the dissimilarity through the non-normalized Victor-Purpura spike distance with a cost parameter of 50 ms (Victor, 2005). We only included those inferred syllable onsets that occurred within the duration of the presented sentence, but not those that occurred before the start of the sentence or after it had ended.

For each simulation, the performance of the network was compared to predictions obtained from a simple constant rhythm. The rate of the constant rhythm for this control model was matched to the frequency of the syllable predictions that the theta network generated during the presentation of a sentence. The onset of the constant rhythm was randomly chosen from the same range of 380-550 ms, as in the case of model simulations. The performance of the syllable predictions achieved by this constant rhythm was quantified through the dissimilarity of the rhythmic predictions from the actual syllable boundaries, in the same way as for actual syllable predictions. Because the syllable prediction generated by this constant rhythm were not influenced by the simulated speech stimulus, they served to estimate the chance-level performance of predicting the syllable onsets.

A non-dimensional parsing score was then computed by subtracting the distance of the inferred to the actual syllable onsets from the analogous measure achieved by the constant rhythm. A parsing score of 0 accordingly reflected no difference from the prediction performance of the constant rhythm, whereas a positive score indicated a prediction of the syllable onsets from the model that was better than in the control. As an additional control measure, we assessed the syllable parsing when only babble noise was presented to the network, with the actual sentence removed from the acoustic stimulus. The obtained parsing scores provided an additional empirical estimate of the chance level.

2.8. Syllable decoding

The excitatory gamma neurons *Ge* received acoustic input that was pre-processed through a model of the auditory periphery, which decomposed the sound into different frequency bands. The activity of the *Ge* neurons therefore partly reflected the spectrotemporal information in the incoming sound. We investigated how well the neural activity encoded the identity of a syllable.

To this end we simulated the neural network response to speech in quiet, as well as to speech in background noise, at various SNRs. We segmented the obtained neural data into subsequent chunks, according to the syllable onsets as inferred from the theta activity described above. Each chunk was assigned the identity of that syllable in the presented sentence during which the corresponding onset was inferred. Moreover, the neural activity in each chunk was characterized by a matrix of pairwise spike distances, for which we employed the non-normalized Victor-Purpura distance with a cost parameter of 60 ms (Victor, 2005) (Fig. 3A). As a result, each single syllable encoded by the model (Fig. 1B, red boxes) was characterized by a particular dissimilarity matrix.

Decoding the identity of a syllable then meant to infer the syllable identity assigned to the chunk from its pairwise spike distance matrix. We performed this decoding in two steps. First, we established the neural responses to speech in quiet as the reference neural activity. For each syllable, this reference neural activity was computed by averaging the pairwise spike distances from all chunks of neural data that were associated to that particular syllable (Fig. 3B).

Second, we employed a nearest centroid algorithm to decode the identity of a syllable associated with a particular chunk of neural data, which could correspond to speech in noise. The reference pairwise spike distances thereby served as centroids. A chunk of neural data was thus assigned that syllable identity to whose reference pairwise spike distance its own pairwise spike distance was closest to. The distance between two matrices of pairwise spike distance was computed as the root mean square of their difference.

2.9. Determining the syllable decoding accuracy

We measured the accuracy of the syllable decoding from the output of the neural network for speech in various levels of background noise, at 11 different SNRs. To this end we performed a large number of trials, in each of which we sought to decode the identity of certain syllables from the network response to speech in noise, using the network response to speech in quiet as a reference.

Neural responses to the speech material were computed as described in Section 2.3. In each classification trial, we chose a random subset of ten syllables (classes) amongst which the neural data was to be classified. For each of the ten syllables we gathered all the neural network's responses to that syllable in a given sentence spoken by a particular speaker, at a particular SNR (testing data) as well as without background noise (training data). For each of the ten syllables, we obtained 100 chunks of corresponding neural data, each characterized by its own dissimilarity matrix.

However, due to inaccuracies in the syllable parsing by the PIN-TH module, the chunks of neural data associated to a particular syllable were sometimes more than 100 and sometimes less. In particular, such deviations are expected for shorter syllables or faster speech production rates (Ghitza, 2011; Hyafil et al., 2015). To balance the classification problem and to prevent biases, in the former case, we selected a random

subset of 100 neural data chunks. In the latter case we selected another subset of 10 syllable labels to be classified, until 100 associated neural data chunks were found for each syllable in the classification trial.

The neural data associated with each syllable, from presenting the sentences in quiet, was then used to establish the reference neural activity. Each chunk of neural data from stimulations employing speech in noise was classified according to the nearest centroid as described above. These predictions were subsequently compared to the actual syllable identities and were averaged to determine the classification accuracy in the decoding trail. Due to the ten different syllables (i.e. classes) that were considered in each trial, the chance level accuracy was 10%.

We performed 200 of such 10-way syllable decoding trials for each of the 11 SNRs for which we simulated the neural network response. The subset of 10 syllables to be classified was chosen at random in each of the 200 trials, but was then kept for each of the SNRs to enable fair comparison between the corresponding syllable decoding accuracies.

2.10. Analysis of the effect of SNR on the speech encoding

The dependency of the syllable decoding accuracy A on the different SNRs could be modelled using a four-parameter sigmoid function:

$$A = \frac{A_{max} - A_{min}}{1 + e^{-k(SNR - SNR_0)}} + A_{min}, \tag{8}$$

in which A_{min} is the minimal decoding accuracy, achieved for a very small SNR, and A_{max} is the maximal decoding accuracy, resulting from a very high SNR. SNR_0 is the SNR at which the decoding accuracy is the average of the maximal and the minimal value, that is, the SNR at which the decoding accuracy is halfway between A_{min} and A_{max}. SNR_0 may therefore be related to the 50% speech reception threshold (SRT) that is commonly used to quantify the level of speech-in-noise comprehension in behavioural experiments. k determines the slope of the curve at SNR_0.

To obtain the model parameters of Eq. (8), as well as their confidence intervals, we employed a bootstrapping procedure (Davison and Hinkley, 1997). The 200 trials of syllable decoding, performed for the eleven different SNRs, resulted in 2,200 datapoints. We resampled these 10,000 times with replacement, and each time computed the parameters of the sigmoidal fit through non-linear least squares (Levenberg-Marquardt algorithm (Marquardt, 1963)). We thereby obtained empirical distributions for each model parameter. The mean value of each parameter followed as the mean of the corresponding distribution, and the associated (100-n)% confidence interval was computed as the range between the distribution's $(\frac{n}{2})^{\text{th}}$ and the $(100 - \frac{n}{2})^{\text{th}}$ percentile. The optimal curve fitted to the data and its confidence bands were computed from these values.

We modelled the effect of background noise on the syllable parsing score through a sigmoidal function as well. The parameters of the sigmoidal fit and their confidence intervals were determined analogously to dependence of the syllable decoding accuracy on the SNR set out above.

2.10. Quantifying the contributions of spectral cues to the speech encoding in the model

To identify the contributions of spectral cues to the syllable parsing and encoding in the model, we repeated the simulations of speech in background noise, but with randomly shuffled auditory channels. Specifically, for each simulation of the model, the 32 auditory channels that contained the auditory input were randomly re-ordered. The time course of each channel remained unchanged, so that the net acoustic input to the model remained the same as for the original stimulus.

The shuffled acoustic inputs were then processed in the model as specified in Section 2.4. In particular, the randomly shuffled auditory channels were projected to the *Ge* neurons and to the population of relay neurons, which provided input to the population of *Te* neurons. The model simulations employed the same sentences and SNRs as in the previous experiment (see Section 2.3 for details). Syllable parsing and decoding were analysed as described in Sections 2.7–2.10. In particular, the model simulations of the original, unshuffled, clean sentences were used as a reference to evaluate the syllable decoding accuracy of the shuffled input.

2.11. Modelling the effects of external electrical stimulation on the speech encoding

To investigate the effects of external electrical stimulation on the encoding of speech in the model, we ran the same model simulations as specified in Section 2.3, but this time simulating the application of external alternating current as well. The stimulation waveforms used and their alignment with respect to the acoustic input were specified in Section 2.5.

The analysis of the syllable parsing and decoding was the same as described in Section 2.7–2.10. Importantly, for syllable decoding, the stimulation waveform was applied also when speech without background noise was simulated in the model. This meant that the centroids of the syllable classifier were computed from speech in quiet, but with added current stimulation. We chose this approach because the neural encoding of speech, including speech in quiet, was likely affected by the applied current. Our goal was, however, to assess the impact of current stimulation on the network's encoding of speech in noise, and not on speech in quiet. We therefore employed the neural responses to speech in quiet during current stimulation as a reference to assess how the applied stimulation influences the consistency of the neural code across SNRs for a given type of stimulation.

3. Results

3.1. Intrinsic network activity

The PIN-G and the PIN-TH modules in the network generated self-sustained rhythmic activity in the gamma (25-40 Hz) and in the theta (4-8 Hz) frequency range, respectively (Fig. 1B). Through the unidirectional coupling from the *Te* to the *Ge* neurons (Fig. 1A, red), the theta rhythm modulated the faster gamma activity. In particular, each burst of spikes generated in the PIN-TH module reset the phase of the faster gamma oscillations (Fig. 1B). We quantified this coupling through computing the phase-amplitude modulation index between the LFP of the PIN-TH module and the LFP of the PIN-G module, when processing an exemplary sentence preceded by a period of silence and without additional current stimulation (Fig. 1C). We found that the neural activity between 5 - 12 Hz modulated the faster activity in the gamma band, above 25 Hz.

3.2. The neural network's encoding of speech in noise

When the network was presented with speech, the theta rhythm aligned to the syllable onsets (Fig. 1B). We quantified this alignment by computing a syllable parsing score, and used it to systematically quantify how well the network parsed syllables when speech was presented in different levels of babble noise. To estimate the empirical chance level, we presented the neural network with babble noise alone, and computed the syllable parsing score that would have been associated with the missing speech signal.

For the lowest SNR that we considered, -25 dB, the syllable parsing score attained a very low value of -0.03 ± 0.05 (mean and 95% CI, Fig. 4A, blue). This was comparable to the empirically estimated chance level of 0.01 ± 0.05 (mean and 95% CI, Fig. 4A, red). The syllable parsing at this low SNR was therefore insignificant. However, SNRs of -10 dB or higher led to syllable parsing score that exceeded the chance level. For the highest SNR of 25 dB that we simulated, the score reached 0.88 ± 0.05 (mean and 95% CI).

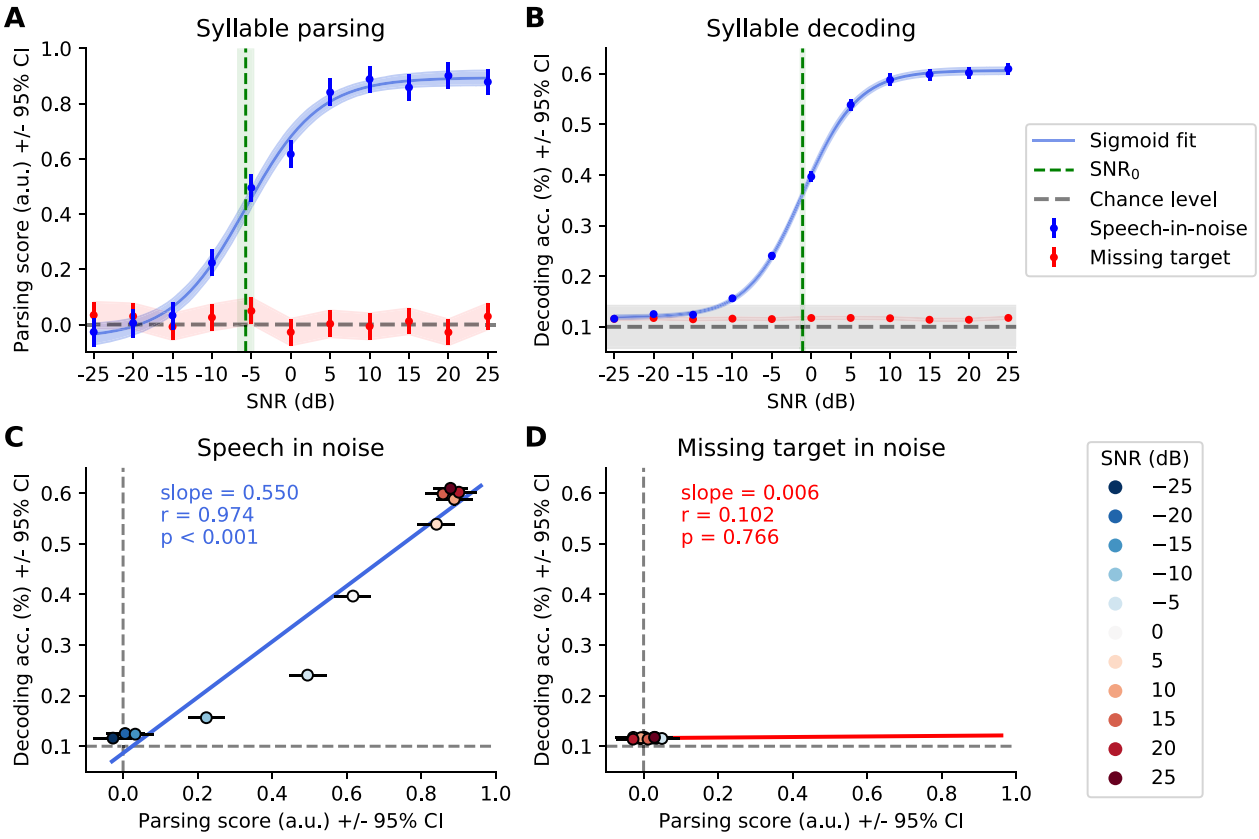

Fig. 4. Speech-in-noise encoding in the model. (A) The syllable parsing by the theta module (blue) was at chance level (grey) for high levels of background noise (low SNR), but exceeded chance level for SNRs above -15 dB. It saturated at a value of around 0.9 for high SNRs, following a sigmoidal relationship with an inflection point at the SNR of -5.7 dB (green). Syllable parsing did not exceed the chance level when the speech signal was absent from the acoustic input (red). **(B)** The accuracy of the syllable decoding (blue) from the neural response of the gamma module exhibited a sigmoidal dependence on the level of the background noise as well. The decoding accuracy was above the chance level (grey) when the SNR was -10 dB or higher. The inflection point of the sigmoidal fit occurred at an SNR of -1.1 dB (green). No significant syllable decoding could be achieved when the speech signal was removed from the background noise (red). **(C)** The syllable decoding accuracy increases monotonously with the syllable parsing score, with increasing SNR. The correlation between the two measures is statistically highly significant ($p = 4 \cdot 10^{-7}$). **(D)** A control computation in which syllable parsing and syllable decoding are obtained from sound mixtures in which the target speech signal has been removed shows performance that is only at the chance level (grey).

To interpret the magnitudes of the parsing scores, we computed the maximal parsing score, which followed from the true syllable onsets. We obtained a maximal parsing score of 9.21 ± 0.02 (mean and 95% CI). Likewise, the parsing score of 0 reflected an insignificant parsing that was equal to that of the null model (Fig. 4A, grey dashed). The maximal parsing scores obtained from the spiking neural network were therefore only about 10% of the maximal possible value, that is, the one that would result from perfect alignment of the predicted and actual syllable onsets.

The dependence of the parsing score on the SNR could be fitted well by a sigmoidal curve (Fig. 4A, blue). The inflection point of the sigmoid, that is, the SNR at which the syllable parsing score was midway between the minimal and the maximal value, occurred at -5.7 dB ± 1.0 dB (mean and 95% CI).

The excitatory neurons of the PIN-G module, the *Ge* neurons, were influenced by the PIN-TH module. At the same time, the *Ge* neurons were stimulated by the sound as well, in a tonotopic fashion (Fig. 1A). While the PIN-TH module could parse syllables, the neuronal activity of the faster PIN-G module could therefore encode the identity of the corresponding syllable. We determined the accuracy of the syllable encoding by assessing how well syllables could be decoded from the spiking activity of the *Ge* neurons.

Because we decoded syllable identities out of ten possible choices, the chance level for the decoding accuracy was 10%. We verified this chance level by assessing the syllable decoding when only background noise was presented to the neural network. This yielded a decoding accuracy of 11.6% ± 0.3% (mean and 95% CI), approximately in line with the chance level (Fig. 4B, red).

We found that the accuracy of the decoding of syllables in background noise, as a function of the SNR, followed a sigmoidal curve (Fig. 4B, blue). For the lowest considered SNR of -25 dB, the decoding was poor, with an accuracy of 11.6% ± 0.6% (mean and 95% CI). This low accuracy exceeded the chance level of 10% only slightly.

The largest SNR that we simulated, 25 dB, led, in contrast, to a high decoding accuracy of 61.0% ± 1.1% (mean and 95% CI). Indeed, the decoding accuracy exceeded the chance level already for the comparatively low SNR of -10 dB, as well as for higher SNRs. Fitting a sigmoid to the dependence of the decoding accuracy on SNR showed that the inflection point of the curve was at a SNR of -1.1 ± 0.2 dB (mean and 95% CI).

We also investigated the relationship between syllable parsing scores and syllable decoding accuracies (Fig. 4C). We found a strong positive correlation between the two measures (Pearson's $r = 0.97$, $p < 10^{-6}$). Low parsing scores were accordingly associated to low accuracies of syllable decoding and *vice versa*. The slope of a linear fit was 0.55.

However, the relation between the two scores was not exactly linear. Instead, intermediate SNRs led to relatively higher syllable parsing scores than the syllable decoding accuracies. This behaviour reflected

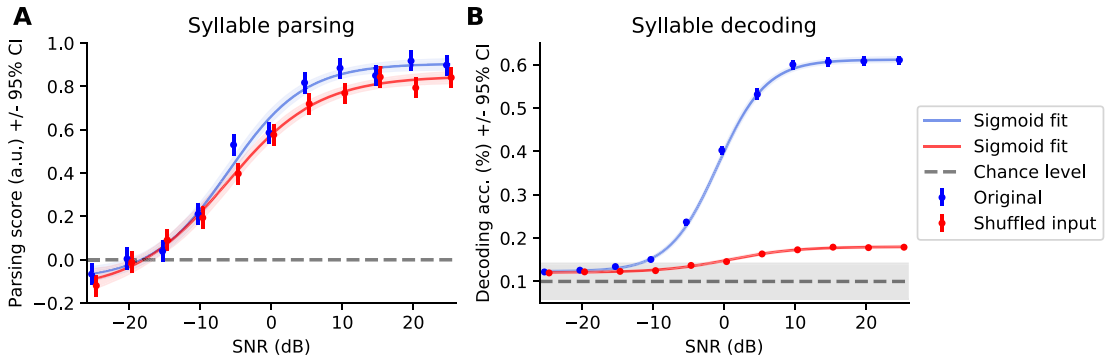

Fig. 5. Encoding of speech with shuffled auditory channels. The figure depicts the syllable parsing scores **(A)** and the decoding accuracies (**B**) obtained for the original acoustic input (blue) or when the auditory channels were randomly shuffled (red). The dashed grey lines represent the chance level for each score, and error bars depict 95% confidence intervals. **(A)** The syllable parsing scores remained at approximately the same level when auditory channels were shuffled, especially for lower SNRs (between -25 and -5 dB). A discrepancy between the shuffled and the original inputs occurred for SNRs above -5 dB. **(B)** In contrast to the syllable parsing, syllable decoding accuracy decreased substantially when the auditory channels were shuffled. In particular, the syllable decoding of the shuffled input remained at or only slightly above chance level for all SNRs. The syllable decoding of the original speech input, however, was significantly higher than the chancel level for SNRs above -10 dB.

our earlier finding that the inflection point of the sigmoidal dependence of the syllable parsing scores on the SNR occurs at a lower SNR, -5.7 dB, than that of the decoding accuracy, -1.1 dB.

As a control, we also assessed the correlation between the syllable parsing score and the decoding accuracy when both were obtained from the background babble noise (Fig. 4D). As expected, the resulting scores were low and not significantly correlated (Pearson's $r = 0.1$, $p = 0.8$).

3.3. Quantifying the contributions of spectral cues to the speech encoding in the model

To investigate the contributions of frequency-specific cues to the model's speech encoding, we shuffled the auditory channels of the acoustic inputs in model simulations. We compared the obtained syllable parsing scores and decoding accuracies with the case when the network was encoding the original acoustic input (Fig. 5).

Shuffling the auditory channels influenced the syllable parsing and decoding differently (Fig. 5). Syllable parsing was not affected strongly by the shuffling, and its dependence on the SNR of the spectrally-shuffled input was comparable to that of the original acoustic signal (Fig. 5A). In particular, for low SNRs, below -5 dB, the results were almost identical. For SNRs above -5 dB, shuffling of the auditory channels led to a slight decrease in performance. The largest difference in the parsing scores between the shuffled and the original acoustic input, a difference of 0.092 a.u., was observed for a SNR of approximately 5 dB . For a SNR of 25 dB, the parsing scores from the two conditions remained different, but the discrepancy between them was smaller (0.063 a.u.).

For syllable decoding, however, the shuffling of auditory channels led to a major deterioration of the classification accuracy (Fig. 5B). Similarly to the syllable parsing scores, for the very low SNRs below -10 dB, the decoding accuracy for both the shuffled and the original input was similar and did not exceed chance level. For SNRs above -10 dB, the results obtained from the two types of input started to diverge. Notably, the syllable decoding accuracy for the shuffled input (Fig. 5B, red) did not exceed the chance level below approximately 0 dB SNR. Even at a SNR of 25 dB it remained substantially below that of the original, non-shuffled, input, reaching only 18.0% ± 0.5% accuracy (mean and 95% CI).

3.4. The effects of the external current stimulation on speech processing in the model

We assessed the effects of the external current stimulation with the speech envelope on the network's encoding of speech stimuli. We investigated three main types of current waveforms: one type that was based on the broad-band speech envelope, a second type that was based on the delta-band portion, and a third type that was based on the theta-band portion of the speech envelope (Fig. 2). For each of these three types, we then considered six different phase shifts. Because the waveforms of each type encompassed more than a single frequency, these phase shifts differed from temporal delays.

In addition, we considered eleven time delays that ranged from -250 ms to 250 ms with 50 ms step. Positive time lags thereby meant that the stimulation onset preceded the sentence that was presented to the model. The phase of the time-shifted waveforms was not manipulated, such that their phase shift was 0°.

Each waveform was applied at three different intensities of 0.1 pA, 0.2 pA and 0.5 pA.

3.4.1. The effects of the external current stimulation on syllable parsing

To quantify the influence of the applied stimulation waveforms on the syllable parsing, governed by the slower theta rhythm in the model, we assessed the obtained parsing scores at the SNR of 0 dB. For this SNR, model simulations without additional current stimulation yielded a parsing score of 0.65 ± 0.05 (mean and 95% CI) (Fig. 4B).

For each applied stimulation waveform, we obtained the parsing score at 0 dB SNR and compared them with the case when no stimulation was applied to the model. The Wilcoxon signed rank test (Wilcoxon, 1945) was used to assess whether the difference between the two was significant. We then applied the Benjamini-Yekutieli correction for false discoveries from multiple comparisons to the obtained p-values (Benjamini and Yekutieli, 2001). The significance threshold for hypothesis testing was set to $p = 0.05$.

For the stimulation waveforms shifted in phase, the theta-band stimulation provided the largest difference to the case without stimulation, consistently across all considered stimulation intensities (Fig. 6A, C, E). The theta-band stimulation notably outperformed the other stimulation waveforms for the phase shifts of 180° and 240°, and provided the largest improvement of the parsing scores. Delta-band stimulation yielded slightly larger improvement than the broad-band waveform for the phase shifts of 180° and 240°. Overall, the effects of the applied stimulation provided phase-dependent modulation, which remained consistent across stimulation intensities. For all types of stimulation waveforms phase shifts ranging from 0° to 120° typically led to the decrease in the parsing scores. In turn, phase-shifts ranging from 180° to 300° facilitated the syllable parsing. While the phase-dependent modulation was observed for all the waveforms, the strength

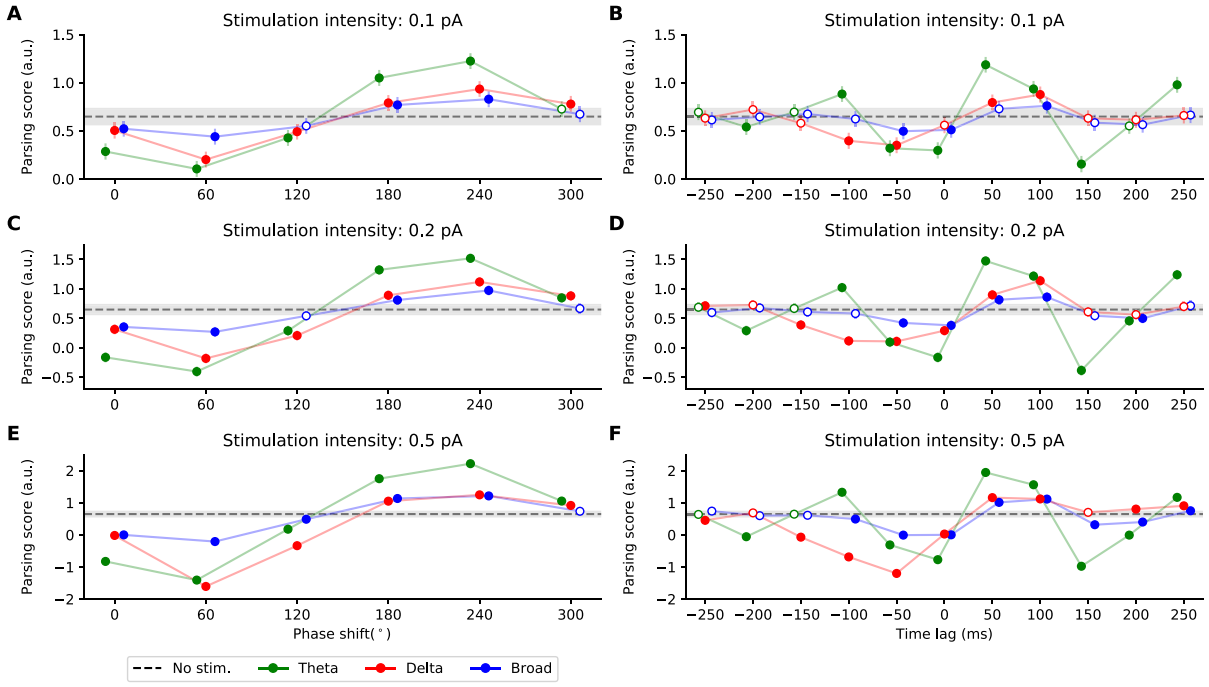

Fig. 6. The effects of the external current stimulation on the syllable parsing. The syllable parsing scores during current stimulation were computed for speech in background noise, at a SNR of 0 dB. We computed the syllable parsing scores for broad-band stimulation (blue), delta-band stimulation (red), and theta-band stimulation (green), and compared the results to the case of no stimulation (black dashed line). The stimulation waveforms were either shifted in phase (**A, C, E**) or in time (**B, D, F**) with respect to the acoustic input. Positive time lags represent stimulation onset preceding the neural processing. Each row of panels shows results for a different stimulation intensity, and the error bars and shaded areas represent 95% confidence intervals. Parsing scores that differ significantly from those obtained without stimulation are indicated by coloured disks ($p < 0.05$, FDR correction for multiple comparisons). Stimulation at phase shifts of about 240° as well as at time shifts of about 50 ms typically enhance the syllable parsing, whereas phase shifts of 60° as well as time shifts of about -50 ms lead to a worsening of the syllable parsing.

of the modulation varied depending on the frequency of the stimulation waveform.

For the stimulation waveforms shifted in time, the effect on the syllable parsing depended on the frequency of the stimulation waveform (Fig. 6B, D, F). For theta-band stimulation, the largest improvement was observed for a delay of 50 ms, that is, when the onset of the stimulation preceded the onset of the acoustic input by 50 ms. Additional significant improvements in parsing scores were observed for time lags of -100 ms, 100 ms and 250 ms. Interestingly, the difference between the two pairs of beneficial lags was 150 ms, corresponding to a frequency of approximately 6.67 Hz. In turn, the negative effects of the theta-band stimulation on the parsing scores were observed for -200, -50, 0 and 150 ms. As for the beneficial time lags, the difference between two successive delays were therefore 150 ms as well.

For the delta-band stimulation, the time lag that led to the largest improvement of the parsing score was between 50 and 100 ms, depending on the stimulation intensity. Similarly to the stimulation with different phase shifts, the effects of the delta-band stimulation at the best time lag were smaller than for the theta-band stimulation, but were larger than the broadband stimulation, across all stimulation intensities. Delta-band stimulation that proceded the acoustic input, that is, at negative time lags, led to a decrease of the parsing scores. The size of this decrease depended on the stimulation intensity and was comparable to that of the theta-band stimulation.

Stimulation with the broadband waveforms shifted in time influenced syllable parsing the least. Notably, only the time lag of 100 ms led to a consistent improvement of parsing scores across all three stimulation intensities. For the two higher stimulation intensities, the time lags of 50 ms (at 0.2 pA, 0.5 pA) and 250 ms (at 0.5 pA) also facilitated syllable parsing, but not as strongly as at the delay of 100 ms.

3.4.2. *The effects of the external current stimulation on syllable decoding*

To assess the neural network's speech encoding during stimulation, we measured the syllable decoding accuracies at the SNR of -1.1 dB. This SNR yielded, without current stimulation, a decoding accuracy of 36.4% ± 0.7% (mean and 95% CI) that was halfway between the minimal and the maximal accuracy (Fig. 4B).

For each type of stimulation waveform, we then established whether the obtained syllable decoding accuracy was significantly different from the one that resulted in the absence of current stimulation. To this end, we obtained the empirical distribution of the syllable decoding accuracies without current stimulation at -1.1 dB SNR through a bootstrapping procedure as described in Section 2.10. This empirical distribution represented the lack of the effects of the applied stimulation. For each stimulation waveform, it was then compared to the syllable decoding accuracy that resulted when current stimulation was applied, to establish an empirical p-value (two tailed) for the obtained decoding accuracy. We then applied the Benjamini-Yekutieli correction for false discoveries from multiple comparisons to the obtained p-values. The significance threshold for hypothesis testing was set to $p = 0.05$.

The lowest stimulation intensity that we considered was 0.1 pA, leading to a depolarization of the membrane potential of a stimulated isolated neuron by 1 mV. Stimulation at such a low intensity led only to significant change in the decoding accuracy for the delta-band stimulation (Fig. 7A, B). In particular, delta-band stimulation at certain phase shifts and time shifts worsened the syllable decoding: a phase shift of 180° resulted in a lower decoding accuracy of 35.0% ± 0.7%, a phase shift of 240° reduced the decoding accuracy to 35.2% ± 0.7%, a time lag of -150 ms yielded a decoding accuracy of 34.9% ± 0.6%, and a time shift of 250 ms lowered the decoding accuracy to 35.2% ± 0.7%.

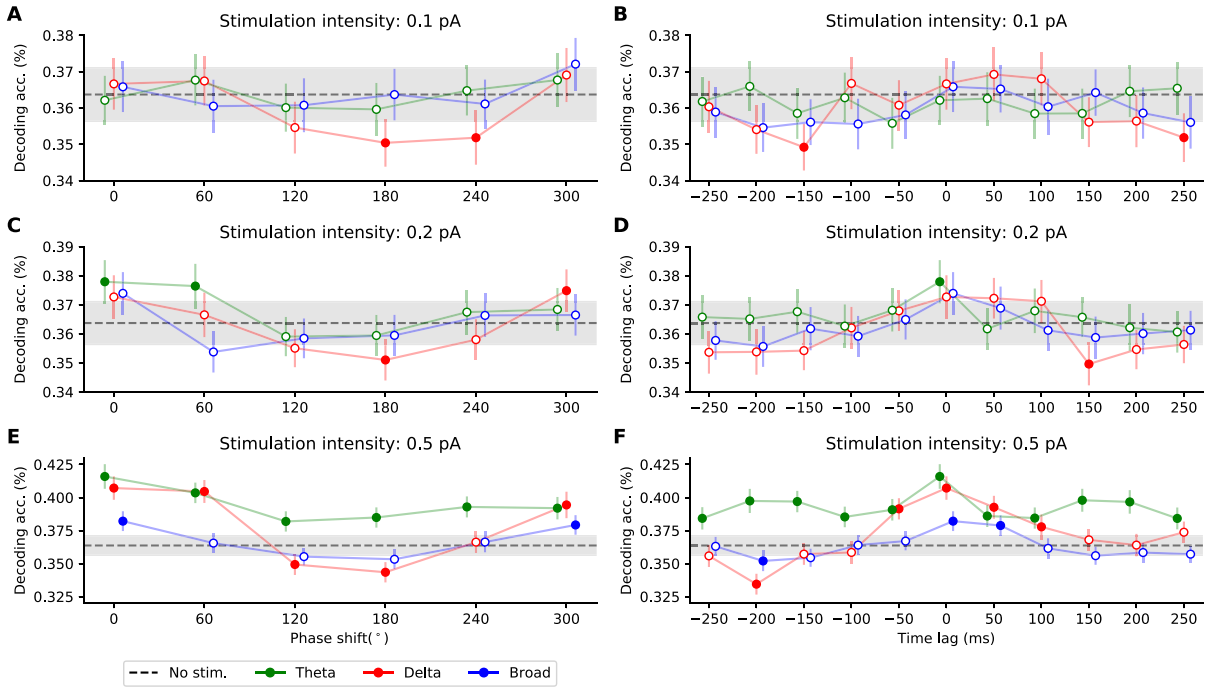

Fig. 7. The effects of the external current stimulation on the syllable encoding. Syllable decoding accuracies during current stimulation were computed for speech in background noise, at a SNR of -1.1 dB. We computed the syllable decoding accuracies for broad-band stimulation (blue), delta-band stimulation (red), and theta-band stimulation (green), and compared the results to the case when no stimulation was applied to the model (black dashed line). The stimulation waveforms were either shifted in phase (**A, C, E**) or in time (**B, D, F**). Positive time lags represent stimulation waveform preceding the acoustic signal. The error bars and shaded areas represent the 95% confidence intervals. Decoding accuracies that differed significantly from the case when no stimulation was applied to the model are indicated by coloured disks ($p < 0.05$, FDR correction for multiple comparisons).

We then investigated an intermediate stimulation intensity of 0.2 pA, leading to a 2-mV depolarization of an isolated neuron. This led to significant differences in the syllable decoding accuracy, as compared to no stimulation, for theta- and delta-band stimulation, but not for the broad-band current waveforms (Fig. 7C, D). In particular, theta-band stimulation yielded significant improvement for phase shifts of 0°, resulting in a syllable decoding accuracy of 37.8% ± 0.7%, and 60°, increasing the accuracy to 37.8% ± 0.8%, as well as at a 0 ms time lag, yielding a decoding accuracy of 37.8% ± 0.7%. In turn, delta-band stimulation yielded significant improvement only for 300° phase shift, increasing the decoding accuracy to 37.5% ± 0.7%. It also led to a significant decrease in decoding accuracy, namely for a 180° phase shift that yielded an accuracy of 35.3% ± 0.7% as well as for a 150 ms time lag that lowered the accuracy to 34.9% ± 0.7%.

At the highest considered stimulation intensity of 0.5 pA, the narrow-band stimulation waveforms had, at most phase shifts, a significant impact on the network's speech coding, while the broadband stimulation yielded significant effects only for a couple of phase and time shifts (Fig. 7E, F). The largest improvement of about 5% in the syllable decoding accuracy emerged for the theta-band stimulation aligned with the sentence onset, that is, without a shift in phase or time. This improvement was slightly larger than that observed for the delta-band stimulation without phase- or time shift, and was substantially larger than that resulting from the broad-band stimulation.

4. Discussion

We investigated the influence of alternating current stimulation with the speech envelope on the neural processing of speech in background noise, using a computational model of a spiking neural network. We characterized the network's speech encoding through two measures, the syllable parsing score and the accuracy with which the syllable identity could be decoded from the neural activity. We found that both measures increased with increasing SNR, following a sigmoidal curve. This behaviour resembled psychometric curves of speech comprehension measured behaviourally (Plomp and Mimpen, 1979; Nilsson et al., 1994; Spyridakou et al., 2020). An important characteristic of each sigmoidal curve was the inflection point, that is, the SNR at which the corresponding measure—the syllable parsing score or the syllable decoding accuracy—was midway between the lowest and the highest value. This inflection point occurred, for the syllable decoding accuracy, at an SNR of -1.1 dB (Fig. 4B). At this SNR, the human comprehension of speech in babble noise is about 50%, suggesting that the neural network's speech encoding may capture certain aspects of the neural mechanisms through which humans understand speech in noise.

The inflection point of the sigmoidal curve occurred at a lower SNR of -5.7 dB for the syllable parsing scores, suggesting that syllable parsing was somewhat more robust in the presence of background noise than syllable decoding. Our further investigation employing spectrally-shuffled versions of the acoustic input showed that the syllable parsing depended mainly on the slow amplitude fluctuations in the acoustic input rather than on frequency-specific features (Fig. 5A). On the contrary, the accuracy in the syllable decoding task deteriorated almost completely when the frequency channels of the acoustic input were randomly shuffled (Fig. 5B). Syllable decoding therefore relied mostly on the frequency-specific information in the acoustic input. These differences between the syllable parsing and the syllable decoding highlight the two distinct mechanisms by which both tasks are accomplished in the model, either through the neural output of the PIN-TH module, or through that of the PIN-G network (Fig. 1).

We investigated the effects of the stimulation waveforms that were derived from the speech envelope on the speech encoding in the model. The stimulation waveforms were either narrow- or broadband, and were simulated with different stimulation intensities. In order to quantify the impact of the alignment of the stimulation waveform and the acoustic input, we shifted the stimulation waveform in either phase or time.

With increasing stimulation intensity, the effects on syllable parsing and syllable decoding increased as well (Figs. 6, 7). For syllable parsing, the phase-shifts of the stimulation waveforms led to a modulation pattern that was periodic in the phase. Phase shifts between 0-120° led to a decrease of parsing scores, and phase shifts of 180-300° improved them (Fig. 6A, C, E). Importantly, this behaviour was consistent for the delta-band, theta-band and broadband stimulation, although the size of the effect depended on the frequency band. Specifically, we observed the largest phase-dependent modulation of syllable parsing for the theta-band stimulation, a moderate one for the delta-band waveforms, and the smallest one for the broadband type stimulation.

These findings parallel recent experimental results on a phase-dependent modulation of speech-in-noise comprehension through transcranial current stimulation with the speech envelope (Riecke et al., 2018; Zoefel et al., 2018; Kadir et al., 2020; Keshavarzi et al., 2020). These experiments reported significant effects of tACS on speech comprehension only for theta-band stimulation, which may correspond to our finding of the theta-band stimulation having the largest effect on the speech encoding in the model (Keshavarzi et al., 2020).

In contrast, time shifts of the stimulation waveform led to modulations of the syllable parsing that depended on the frequency of the stimulation waveform (Fig. 6B, D, F). All of the considered stimulation waveforms yielded improved syllable parsing scores when the stimulation preceded the acoustic input by 50 ms to 100 ms. This time lag matches the neural delay of the early cortical processing of speech (Kubanek et al., 2013; Brodbeck and Simon, 2020).

To interpret our results regarding the temporal delays, it is important to note that our model did not include a temporal delay between the acoustic input and the neuronal modules, with the exception of the relay neurons that fed into the theta network and that had delays between 0 and 50 ms. However, the neural responses in the human auditory cortex exhibit delays between approximately 20 ms and hundreds of ms (Pickles, 2013). In particular, the primary auditory cortex responds largely at delays between 50 to 100 ms. If we assume that our model of a spiking neural network corresponds to a part of the primary auditory cortex, we therefore need to account for an additional delay of 50 ms to 100 ms in the neural response with respect to the acoustic signal. The maximal enhancement of the syllable parsing score would then occur for neurostimulation waveforms that had no time shift with respect to the acoustic signal. This interpretation of our model prediction agrees with recent behavioural experiments that found that theta-band stimulation with no temporal delay led to the largest improvement in speech comprehension (Keshavarzi and Reichenbach, 2020; Keshavarzi et al., 2020).

The magnitude of the enhancement of syllable parsing at the time shifts of 50 ms to 100 ms depended on the type of stimulation waveform. Consistent with the results obtained for the phase-shifted waveforms, the largest improvement was obtained for theta-band stimulation, followed by its delta-band counterpart, while broadband stimulation yielded the smallest improvements.

Notably, only theta-band stimulation waveform led to substantial improvement of parsing scores for the two further temporal delays, at approximately -100 ms and at approximately 250 ms. Both time lags differ from the time lags that led to the highest syllable parsing score, 50 ms to 100 ms, by 150 ms to 200 ms. This apparent periodicity in the modulation of the syllable parsing score at a period of 150 ms to 200 ms may reflect the periodicity in the theta-band waveform. The period of 150 ms to 200 ms corresponds indeed to a frequency between 5 Hz and 7 Hz, so entirely in the theta frequency range.

Our observation of theta-band stimulation yielding the largest enhancement of syllable parsing presumably reflects the fact that the theta-band stimulation had a frequency range similar to the intrinsic activity of the PIN-TH network, about 5-10 Hz. The theta-band stimulation could therefore efficiently entrain the oscillations in the theta module on the per-cycle basis (Herrmann et al., 2016). Delta- and broadband stimulation waveforms could entrain theta oscillations as well, and consequently influence syllable parsing, but to a smaller extent, especially regarding an enhancement. For the delta-band stimulation, this was likely due to the mismatch between the frequency bands: the delta-band frequencies, at 1 - 4 Hz, were subharmonics of those that occurred in the intrinsic activity of the PIN-TH module. As a result, one cycle of the applied stimulation affected, on average, two cycles of the theta-band oscillations tracking syllable onsets, leading to a weaker entrainment and associated improvement in syllable parsing. Because the broadband stimulation included both the theta and the delta band, it presumably led to interferences between the two and therefore to a further weakening of the effect. Moreover, the delta- and theta-band waveforms but not the broadband waveforms had been processed so that their maxima and minima occurred at the same values, which might have further increased the efficacy of the delta- and theta stimulation.

Syllable decoding was overall affected by the current stimulation to a lesser degree than syllable parsing (Fig. 7). In particular, the lower stimulation intensities of 0.1 pA and 0.2 pA yielded barely a significant modulation of the syllable decoding. At the highest stimulation intensity of 0.5 pA, the effect depended on the phase and time shifts. The largest improvement in the syllable decoding accuracy was achieved when the applied waveform was aligned to the speech signal, without an additional phase shift, whereas the opposite phase shift of 180° yielded the worst syllable decoding accuracy. This parallels recent experimental findings that have found the phase-dependent modulation of speech-in-noise comprehension due to current stimulation with the significant improvement for a phase shift of 0° and the worst performance obtained for a phase shift of 180° (Keshavarzi and Reichenbach, 2020; Keshavarzi et al., 2020).

Somewhat unexpectedly, at the highest stimulation intensity, theta-band stimulation consistently improved the decoding accuracy for all the phase and time lags that we considered. This result presumably reflected the matching of the theta stimulation to the intrinsic rhythm of the theta module that parsed the syllables. Because the syllable decoder was established using the model's response to clean sentences under a certain type of stimulation, the decoding scheme emphasized the consistency of the neural code across SNRs under certain stimulation condition. Since the effects of the theta-band stimulation on the parsing of syllables in the model were overall the strongest, the encoding of speech could therefore benefit from theta-band stimulation, for different time and phase shifts. However, no such effect was observed for the delta- and broadband stimulation waveforms, whose influence on the syllable parsing was notably weaker.

The syllable decoding under the strongest theta-band stimulation depended nonetheless on both phase and time shifts. The largest enhancement of the decoding accuracy was obtained in the absence of either phase or time shifts, that is, at shifts of 0° and 0 ms. Regarding phase shifts, the worst performance was obtained when the waveform was shifted by 120 ° - 240°. Regarding time shifts, the performance decreased symmetrically for both negative and positive shifts up to ± 100 ms, and then increased again towards peaks between ± 150 ms to ± 200 ms. The emergence of these peaks that differed from the largest peak at 0 ms by 150 ms to 200 ms was reminiscent of the dependence of the syllable parsing score on the time shifts, for the theta-band stimulation (Fig. 6F). As in that case, the dependence of the syllable encoding on the time lags likely reflected periodicity in the intrinsic rate of the syllable-parsing PIN-TH module, between 5-10 Hz, yielding a period between 100 ms to 200 ms.

For the strongest stimulation intensity of 0.5 pA, and without phase or time lag, the delta-band stimulation yielded an enhancement of the syllable decoding accuracy that was only slightly below that of the theta stimulation. However, as opposed to the stimulation with the theta-band portion of the speech envelope, the delta-band stimulation could also decrease the accuracy, such as at phase shifts of 120° and of 180° as well as at a time lag of -200 ms. The significant decrease in syllable decoding at these phase and time lags reflects a substantial deterioration of the consistency of the neural code across SNRs during those stimulations.

Broadband stimulation at the intensity of 0.5 pA yielded substantially smaller effects on syllable decoding accuracy than both the delta- and the theta-band stimulation. We found significant improvements of the accuracy only for the phase shifts of 0° and of 300°, as well as for time shifts of 0 ms and 50 ms. Similar to the delta-band stimulation, the only significant negative effect was observed for a time lag of -200 ms, although the effect was smaller.

Surprisingly, the best and worst stimulation phases and time lags for the syllable parsing differed from those of the syllable decoding accuracies. The best syllable parsing was obtained for a phase shift of 240°, yielding a phase advance of 120° with respect to the unshifted waveform. We obtained similarly the worst syllable parsing at a phase shift of 60°, also at a phase advance of 120° as compared to the phase at which the worst syllable decoding accuracy occurred. Regarding the time delays, the largest improvement in the parsing scores were obtained for either 50 ms (theta-band stimulation) or for 100 ms (delta- and broadband stimulation). The best syllable decoding resulted in the absence of a time delay.

These systematic phase and time differences between the influence of the current stimulation on the syllable parsing and on the syllable decoding were unexpected. They reiterate that the parsing of syllables and the encoding of the syllable content in the network activity were governed by distinct mechanisms, implemented by the two modules constituting the network (Fig. 1). First, the activity in each module was likely influenced by the external current in a different way. In particular, the frequency of the theta rhythm was similar to that of the exogenous, envelope-shaped, stimulation waveform, and could therefore be entrained by the latter (Herrmann et al., 2016). The gamma activity, in contrast, had higher intrinsic frequency and therefore the substantially slower external alternating current stimulation waveform could only temporally modulate, rather than directly entrain, the activity of neurons making up the PIN-G network (Fröhlich and McCormick, 2010).

Second, the influence of the current stimulation on the two tasks of syllable parsing and syllable decoding differed as well. In particular, the speech envelope reflects mainly the voiced parts of speech, which generally have a larger amplitude than the voiceless parts (Grant et al., 1985; Shannon et al., 1995; Biesmans et al., 2017). Syllables begin, however, often with a voiceless part. Even in syllables beginning with a voiced part, their onsets precede the majority of their energetic content. The speech envelope, shifted to have a phase or time advance, therefore aligns better with the syllable onsets than the unshifted envelope. A phase or time advance of the current stimulation could accordingly lead to better syllable parsing in the model. In contrast, the syllable decoding from the model output relied mostly on the voiced parts of speech, which yield larger activations of auditory channels than the voiceless parts (Chi et al., 2005). The peaks of the speech envelope aligned with the acoustic input therefore coincided with the stimulus-driven current delivered to the PIN-G module, what in turn facilitated the syllable encoding.

Our model therefore suggests that the stimulation waveforms that are optimal for syllable parsing are not optimal for syllable decoding, and *vice versa*. However, syllable decoding partially depends on syllable parsing. The different influence of the current stimulation on both processes accordingly implies that the neurostimulation's effect on speech encoding in the model is partly inhibited by these interferences.

An important limitation of our model is that it operates only in a feed-forward fashion. The acoustic stimuli and the current stimulation serve as input to the model. The PIN-TH module parses syllables and feeds forward to the PIN-G module, the neural activity of which allows to decode the syllable identity. The brain, in contrast, employs many feedback loops. In particular, attention to one of several acoustic streams as well as linguistic predictions likely act as top-down effects on speech coding (Zion Golumbic et al., 2013; O'Sullivan et al., 2015; Etard and Reichenbach, 2019; Weissbart et al., 2019). Incorporating such higher-level cognitive processes as feedback mechanisms in the model will likely influence the neural network's capability for speech encoding. Importantly, incorporation of these mechanisms in the model will allow us to simulate how they may be influenced by tACS and determine their contribution to the neural processing underlying speech-in-noise comprehension.

Because our model is based on the hypothesis of speech encoding through coupled neural oscillations (Giraud and Poeppel, 2012; Hyafil et al., 2015), it might be used in the future to generate further predictions of how speech processing can be impacted by neurostimulation. Experimental verification or falsification of such predictions may allow to further establish the neural mechanisms of speech processing, and in particular to further investigate the role of coupled cortical oscillations. Moreover, our modelling framework may also be adapted to assess the effects of neurostimulation on the neural processing of other sounds, such as on music perception that may involve coupled oscillations as well.

Further developments of the model may also integrate it with structural modelling that seeks to estimate the current intensity in different brain regions for a certain placement of the electrodes and applied current (Thielscher et al., 2015; Huang et al., 2019). In particular, such modelling may be based on subject-specific data like MRI images that might allow to obtain subject-specific outcomes, for instance regarding current intensities and neural delays. Integrating structural and functional modelling might therefore facilitate the understanding of inter-subject variations as well as allow to optimize stimulation parameters for an individual subject.

Declaration of Competing Interest

The authors declare that the research was conducted in the absence of any conflict of interest.

CRediT authorship contribution statement

Mikolaj Kegler: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Visualization, Validation, Writing - original draft, Writing - review & editing. **Tobias Reichenbach:** Conceptualization, Funding acquisition, Investigation, Supervision, Project administration, Resources, Writing - review & editing.

Funding

This study was supported by the EPSRC Centre for Doctoral Training in Neurotechnology for Life and Health and by EPSRC grant EP/R032602/1 to T.R.

Acknowledgements

We would like to thank Alexandre Hyafil and Anirudh Kulkarni for discussions on the implementation of the model. We would like to thank Shabnam Kadir for helpful discussions, as well as for sharing the audio material used in her experimental study. We are grateful to the Imperial College High-Performance Computing Service (https://doi.org/10.14469/hpc/2232) for supporting this study.

Code & Data availability

All of the above described model and analysis were implemented via custom written Python scripts using SciPy package (Jones et al., 2001), unless stated otherwise. Implementation of the framework introduced here as well as analysis tools are openly available at https://github.com/MKegler/SpeechTACSmodel. TIMIT speech corpus (Garofolo et al., 1993) used in the model simulations is available at https://catalog.ldc.upenn.edu/LDC93S1.

References

Adam V., Hyafil A. (2020) Non-linear regression models for behavioral and neural data analysis. arXiv, 2002.00920.

Ainsworth, M., Lee, S., Cunningham, M.O., Roopun, A.K., Traub, R.D., Kopell, N.J., Whittington, M.A., 2011. Dual γ rhythm generators control interlaminar synchrony in auditory cortex. J. Neurosci. 31 (47), 17040–17051.

Ali, M.M., Sellers, K.K., Fröhlich, F., 2013. Transcranial alternating current stimulation modulates large-scale cortical network activity by network resonance. J. Neurosci. 33 (27), 11262–11275.

Anderson, S., Kraus, N., 2010. Sensory-cognitive interaction in the neural encoding of speech in noise: a review. J. Am. Acad. Audiol. 21 (9), 575–585.

Benjamini, Y., Yekutieli, D., 2001. The control of the false discovery rate in multiple testing under dependency. Ann. Stat. 29 (4), 1165–1188.

Bestmann, S., de Berker, A.O., Bonaiuto, J., 2015. Understanding the behavioural consequences of noninvasive brain stimulation. Trends Cognit. Sci. 19 (1), 13–20.

Biesmans, W., Das, N., Francart, T., Bertrand, A., 2017. Auditory-inspired speech envelope extraction methods for improved EEG-based auditory attention detection in a cocktail party scenario. IEEE Trans. Neural Syst. Rehabil. Eng. 25 (5), 402–412.

Bonaiuto, J.J., Bestmann, S., 2015. Understanding the nonlinear physiological and behavioral effects of tDCS through computational neurostimulation. Progr. Brain Res. 222, 75–103.

Brodbeck, C., Simon, J.Z., 2020. Continuous speech processing. Current Opin. Physiol. 18, 25–31.

Broderick, M.P., Anderson, A.J., di Liberto, G.M., Crosse, M.J., Lalor, E.C., 2018. Electrophysiological correlates of semantic dissimilarity reflect the comprehension of natural, narrative speech. Current Biol. 28 (5), 803–809.

Brosch, M., Budinger, E., Scheich, H., 2002. Stimulus-related gamma oscillations in primate auditory cortex. J. Neurophysiol. 87 (6), 2715–2725.

Cakan, C., Obermayer, K., 2020. Biophysically grounded mean-field models of neural populations under electrical stimulation. PLoS Comput. Biol. 16 (4), e1007822.

Cardin, J.A., Carlén, M., Meletis, K., Knoblich, U., Zhang, F., Deisseroth, K., Tsai, L.H., Moore, C.I., 2009. Driving fast-spiking cells induces gamma rhythm and controls sensory responses. Nature 459 (7247), 663–667.

Chi, T., Ru, P., Shamma, S.A., 2005. Multiresolution spectrotemporal analysis of complex sounds. J. Acoust. Soc. Am. 118 (2), 887–906.

Datta, A., Bansal, V., Diaz, J., Patel, J., Reato, D., Bikson, M., 2009. Gyri-precise head model of transcranial direct current stimulation: improved spatial focality using a ring electrode versus conventional rectangular pad. Brain Stimul. 2 (4), 201–207.

Davison, A.C., Hinkley, D.V., 1997. Bootstrap Methods and Their Application. Cambridge university press.

Ding, N., Melloni, L., Zhang, H., Tian, X., Poeppel, D., 2016. Cortical tracking of hierarchical linguistic structures in connected speech. Nat. Neurosci. 19 (1), 158–164.

Drullmana, R., 1995. Speech intelligibility in noise: Relative contribution of speech elements above and below the noise level. J. Acoust. Soc. Am. 98 (3), 1796–1798.

Etard, O., Reichenbach, T., 2019. Neural speech tracking in the theta and in the delta frequency band differentially encode clarity and comprehension of speech in noise. J. Neurosci. 39 (29), 5750–5759.

Fröhlich, F., 2015. Experiments and models of cortical oscillations as a target for noninvasive brain stimulation. Progr. Brain Res. 222, 41–73.

Fröhlich, F., McCormick, D.A., 2010. Endogenous electric fields may guide neocortical network activity. Neuron 67 (1), 129–143.

Fröhlich, F., Schmidt, S.L., 2013. Rational design of transcranial current stimulation (TCS) through mechanistic insights into cortical network dynamics. Front. Hum. Neurosci. 7, 804.

Fröhlich, F., Sellers, K.K., Cordle, A.L., 2015. Targeting the neurophysiology of cognitive systems with transcranial alternating current stimulation. Expert Rev. Neurotherapeut. 15 (2), 145–167.

Garofolo, J.S., Lamel, L.F., Fisher, W.M., Fiscus, J.G., Pallett, D.S., Dahlgren, N.L., 1993. DARPA TIMIT Acoustic Phonetic Continuous Speech Corpus (LDC93S1). Linguistic Data Consortium, Philadelphia.

Ghitza, O., 2011. Linking speech perception and neurophysiology: speech decoding guided by cascaded oscillators locked to the input rhythm. Front. Psychol. 2, 130.

Giraud, A.-L., Kleinschmidt, A., Poeppel, D., Lund, T.E., Frackowiak, R.S.J., Laufs, H., 2007. Endogenous cortical rhythms determine cerebral specialization for speech perception and production. Neuron 56 (6), 1127–1134.

Giraud, A.-L., Poeppel, D., 2012. Cortical oscillations and speech processing: emerging computational principles and operations. Nat. Neurosci. 15 (4), 511–517.

Grant, K.W., Ardell, L.A.H., Kuhl, P.K., Sparks, D.W., 1985. The contribution of fundamental frequency, amplitude envelope, and voicing duration cues to speechreading in normal-hearing subjects. J. Acoust. Soc. Am. 77 (2), 671–677.

Gross, J., Hoogenboom, N., Thut, G., Schyns, P., Panzeri, S., Belin, P., Garrod, S., 2013. Speech rhythms and multiplexed oscillatory sensory coding in the human brain. PLoS Biol. 11 (12), e1001752.

Han, C., O'Sullivan, J., Luo, Y., Herrero, J., Mehta, A.D., Mesgarani, N., 2019. Speaker-independent auditory attention decoding without access to clean speech sources. Sci. Adv. 5 (5), eaav6134.

Helfrich, R.F., Schneider, T.R., Rach, S., Trautmann-Lengsfeld, S.A., Engel, A.K., Herrmann, C.S., 2014. Entrainment of brain oscillations by transcranial alternating current stimulation. Current Biol. 24 (3), 333–339.

Herrmann, C.S., Murray, M.M., Ionta, S., Hutt, A., Lefebvre, J., 2016. Shaping intrinsic neural oscillations with periodic stimulation. J. Neurosci. 36 (19), 5328–5337.

Hickok, G., Poeppel, D., 2007. The cortical organization of speech processing. Nat. Rev. Neurosci. 8 (5), 393–402.

Huang, Y., Datta, A., Bikson, M., Parra, L.C., 2019. Realistic volumetric-approach to simulate transcranial electric stimulation—ROAST—a fully automated open-source pipeline. J. Neural Eng. 16 (5), 56006.

Huang, Y., Parra, L.C., 2019. Can transcranial electric stimulation with multiple electrodes reach deep targets? Brain Stimul. 12 (1), 30–40.

Hutcherson, R.W., Dirks, D.D., Morgan, D.E., 1979. Evaluation of the speech perception in noise (SPIN) test. Otolaryngology–Head Neck Surg. 87 (2), 239–245.

Hyafil, A., Fontolan, L., Kabdebon, C., Gutkin, B., Giraud, A.-L., 2015. Speech encoding by coupled cortical theta and gamma oscillations. eLife 4, e06213.

Iotzov, I., Parra, L.C., 2019. EEG can predict speech intelligibility. J. Neural Eng. 16 (3), 036008.

Jadi, M.P., Sejnowski, T.J., 2014. Cortical oscillations arise from contextual interactions that regulate sparse coding. Proc. Natl. Acad. Sci. 111 (18), 6780–6785.

Jones E., Oliphant T., Peterson P. (2001) SciPy: Open source scientific tools for Python.

Kadir, S., Kaza, C., Weissbart, H., Reichenbach, T., 2020. Modulation of speech-in-noise comprehension through transcranial current stimulation with the phase-shifted speech envelope. IEEE Trans. Neural Syst. Rehabil. Eng. 28 (1), 23–31.

Kasten, F.H., Duecker, K., Maack, M.C., Meiser, A., Herrmann, C.S., 2019. Integrating electric field modeling and neuroimaging to explain inter-individual variability of tACS effects. Nat. Commun. 10 (1), 5427.

Keshavarzi, M., Kegler, M., Kadir, S., Reichenbach, T., 2020. Transcranial alternating current stimulation in the theta band but not in the delta band modulates the comprehension of naturalistic speech in noise. NeuroImage 210, 116557.

Keshavarzi, M., Reichenbach, T., 2020. Transcranial alternating current stimulation with the theta-band portion of the temporally-aligned speech envelope Improves speech-in-noise comprehension. Front. Hum. Neurosci. 14, 187.

Krause, M.R., Vieira, P.G., Csorba, B.A., Pilly, P.K., Pack, C.C., 2019. Transcranial alternating current stimulation entrains single-neuron activity in the primate brain. Proc. Natl. Acad. Sci. USA 116 (12), 5747–5755.

Kubanek, J., Brunner, P., Gunduz, A., Poeppel, D., Schalk, G., 2013. The tracking of speech envelope in the human cortex. PLoS ONE 8 (1), e53398.

Kullback, S., Leibler, R.A., 1951. On information and sufficiency. Ann. Math. Stat. 22 (1), 79–86.

Lalor, E.C., Foxe, J.J., 2010. Neural responses to uninterrupted natural speech can be extracted with precise temporal resolution. Eur. J. Neurosci. 31 (1), 189–193.

Lesenfants, D., Vanthornhout, J., Verschueren, E., Decruy, L., Francart, T., 2019. Predicting individual speech intelligibility from the cortical tracking of acoustic- and phonetic-level speech representations. Hear. Res. 380, 1–9.

Luo, H., Poeppel, D., 2007. Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. Neuron 54 (6), 1001–1010.

Marquardt, D.W., 1963. An algorithm for least-squares estimation of nonlinear parameters. J. Soc. Ind. Appl. Math. 11 (2), 431–441.

Mazzoni, A., Panzeri, S., Logothetis, N.K., Brunel, N., 2008. Encoding of naturalistic stimuli by local field potential spectra in networks of excitatory and inhibitory neurons. PLoS Comput. Biol. 4 (12), e1000239.

Mesgarani, N., David, S., Fritz, J.B., Shamma, S.A., 2014. Mechanisms of noise robust representation of speech in primary auditory cortex. Proc. Natl. Acad. Sci. USA 111 (18), 6792–6797.

Molinaro, N., Lizarazu, M., 2018. Delta(but not theta)-band cortical entrainment involves speech-specific processing. Eur. J. Neurosci. 48 (7), 2642–2650.

Morillon, B., Liégeois-Chauvel, C., Arnal, L.H., Bénar, C.-G., Giraud, A.-L., 2012. Asymmetric function of theta and gamma activity in syllable processing: an intra-cortical study. Front. Psychol. 3, 248.

Negahbani, E., Kasten, F.H., Herrmann, C.S., Fröhlich, F., 2018. Targeting alpha-band oscillations in a cortical model with amplitude-modulated high-frequency transcranial electric stimulation. NeuroImage 173, 3–12.

Nilsson, M., Soli, S.D., Sullivan, J.A., 1994. Development of the Hearing In Noise Test for the measurement of speech reception thresholds in quiet and in noise. J. Acoust. Soc. Am. 95 (2), 1085–1099.

O'Sullivan, J., Chen, Z., Herrero, J., McKhann, G.M., Sheth, S.A., Mehta, A.D., Mesgarani, N., 2017. Neural decoding of attentional selection in multi-speaker environments without access to clean sources. J. Neural Eng. 14 (5), 056001.

O'Sullivan, J.A., Power, A.J., Mesgarani, N., Rajaram, S., Foxe, J.J., Shinn-Cunningham, B.G., Slaney, M., Shamma, S.A., Lalor, E.C., 2015. Attentional selection in a cocktail party environment can be decoded from single-trial EEG. Cereb. Cortex 25 (7), 1697–1706.

Pickles, J., 2013. An Introduction to the Physiology of Hearing. Brill.

Pikovsky, A., Rosenblum, M., Kurths, J.(Jürgen), 2001. S Synchronization: A Universal Concept in Nonlinear Sciences. Cambridge University Press.

Pillow, J.W., Shlens, J., Paninski, L., Sher, A., Litke, A.M., Chichilnisky, E.J., Simoncelli, E.P., 2008. Spatio-temporal correlations and visual signalling in a complete neuronal population. Nature 454 (7207), 995–999.

Plomp, R., Mimpen, A.M., 1979. Improving the reliability of testing the speech reception threshold for sentences. Int. J. Audiol. 18 (1), 43–52.

Radman, T., Ramos, R.L., Brumberg, J.C., Bikson, M., 2009. Role of cortical cell type and morphology in subthreshold and suprathreshold uniform electric field stimulation in vitro. Brain Stimul. 2 (4), 215–228.

Ray, S., Maunsell, J.H.R., 2011. Different origins of gamma rhythm and high-gamma activity in macaque visual cortex. PLoS Biol. 9 (4), e1000610.

Reato, D., Rahman, A., Bikson, M., Parra, L.C., 2010. Low-intensity electrical stimulation affects network dynamics by modulating population rate and spike timing. J. Neurosci. 30 (45), 15067–15079.

Reato, D., Rahman, A., Bikson, M., Parra, L.C., 2013. Effects of weak transcranial alternating current stimulation on brain activity—a review of known mechanisms from animal studies. Front. Hum. Neurosci. 7, 687.

Riecke, L., Formisano, E., Sorger, B., Başkent, D., Gaudrain, E., 2018. Neural entrainment to speech modulates speech intelligibility. Current Biol. 28 (2), 161–169.
Ruhnau, P., Neuling, T., Fuscá, M., Herrmann, C.S., Demarchi, G., Weisz, N., 2016. Eyes wide shut: transcranial alternating current stimulation drives alpha rhythm in a state dependent manner. Sci. Rep. 6 (1), 27138.
Shamir, M., Ghitza, O., Epstein, S., Kopell, N., Latham, P.E., 2009. Representation of time–varying stimuli by a network exhibiting oscillations on a faster time scale. PLoS Comput. Biol. 5 (5), 1000370.
Shamma, S., 1989. Spatial and temporal processing in central auditory networks. Methods Neuronal Model. From Synapses Netw. 247–289.
Shannon, R., Zeng, F.G., Kamath, V., Wygonski, J., Ekelid, M., 1995. Speech recognition with primarily temporal cues. Science 270 (5234), 303–304.
Shi J., Xu Y., Baraniuk R.G. (2014) Sparse bilinear logistic regression. arXiv, 1404.4104.
Sohal, V.S., Zhang, F., Yizhar, O., Deisseroth, K., 2009. Parvalbumin neurons and gamma rhythms enhance cortical circuit performance. Nature 459 (7247), 698–702.
Soli, S.D., Wong, L.L.N., 2008. Assessment of speech intelligibility in noise with the Hearing in Noise Test. Int. J. Audiol. 47 (6), 356–361.
Spyridakou, C., Rosen, S., Dritsakis, G., Bamiou, D.-E., 2020. Adult normative data for the speech in babble (SiB) test. Int. J. Audiol. 59 (1), 33–38.
Thielscher, A., Antunes, A., Saturnino, G.B., 2015. Field modeling for transcranial magnetic stimulation: a useful tool to understand the physiological effects of TMS? In: Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, pp. 222–225.
Tort, A.B.L., Komorowski, R., Eichenbaum, H., Kopell, N., 2010. Measuring phase-amplitude coupling between neuronal oscillations of different frequencies. J. Neurophysiol. 104 (2), 1195–1210.
Vanthornhout, J., Decruy, L., Wouters, J., Simon, J.Z., Francart, T., 2018. Speech intelligibility predicted from neural entrainment of the speech envelope. J. Assoc. Res. Otolaryngol. 19 (2), 181–191.
Victor, J.D., 2005. Spike train metrics. Current Opin. Neurobiol. 15 (5), 585–592.
Weissbart, H., Kandylaki, K.D., Reichenbach, T., 2019. Cortical tracking of surprisal during continuous speech comprehension. J. Cognit. Neurosci. 32 (1), 155–166.
Wilcoxon, F., 1945. Individual Comparisons by Ranking Methods. Biometrics Bull. 1, 80.
Wilsch, A., Neuling, T., Obleser, J., Herrmann, C.S., 2018. Transcranial alternating current stimulation with speech envelopes modulates speech comprehension. NeuroImage 172, 766–774.
Yang, X., Wang, K., Shamma, S.A., 1992. Auditory representations of acoustic signals. IEEE Trans. Inf. Theory 38 (2), 824–839.
Zaehle, T., Rach, S., Herrmann, C.S., 2010. Transcranial alternating current stimulation enhances individual alpha activity in human EEG. PLoS ONE 5 (11), e13766.
Zion Golumbic, E.M., Ding, N., Bickel, S., Lakatos, P., Schevon, C.A., McKhann, G.M., Goodman, R.R., Emerson, R., Mehta, A.D., Simon, J.Z., Poeppel, D., Schroeder, C.E., 2013. Mechanisms underlying selective neuronal tracking of attended speech at a “cocktail party”. Neuron 77 (5), 980–991.
Zoefel, B., Archer-Boyd, A., Davis, M.H., 2018. Phase entrainment of brain oscillations causally modulates neural responses to intelligible speech. Current Biol. 28 (3), 401–408.